

DOCUMENT RESUME

ED 178 066

IR 007 808

AUTHOR Pearson, Margot; And Others
TITLE The Production and Evaluation of Film Material to Assess the Interpersonal Communication Skills of Medical Students. Research and Development Paper No. 53.
INSTITUTION New South Wales Univ., Kensington (Australia). Tertiary Education Research Centre.
PUB DATE Jul 78
NOTE 33p.; For a related document, see IR 007 791 ; Contains some light print
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Evaluation Methods; Instructional Films; *Interpersonal Competence; Medical Education; *Medical Students; Post Testing; Pretests; Professional Education; *Program Evaluation; *Tests; Videotape Recordings

ABSTRACT

The test described was developed to help assess the interpersonal communication skills of medical students in an introductory clinical studies course at the University of New South Wales designed to improve these skills. The stimulus materials consist of filmed vignettes portraying patients expressing strong emotions to doctors. The vignette ends before the doctor replies, and students are asked to respond to the patient either in writing (open-ended test) or by choosing the most appropriate of five given responses (forced-choice test). The pre- and posttests were administered to first year medical students enrolled in the course and to a control group of first year physics students. Responses to both types of test were rated according to a schedule which assessed the ability to respond facilitatively in a patient-doctor interaction. Findings indicated that the scores of the medical students increased significantly while those of the physics students did not, suggesting that the test was sensitive to changes in the students' interpersonal communication skills. The report includes data, a list of 27 references, and a copy of the open-ended response form. (JEG)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *



RESEARCH AND DEVELOPMENT PAPER No. 53

ED178066

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

ISSN 0313-5543

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY.

ISBN 909528 24 1

THE PRODUCTION AND EVALUATION OF FILM MATERIAL TO ASSESS THE INTERPERSONAL COMMUNICATION SKILLS OF MEDICAL STUDENTS

Margot Pearson, T.E.R.C., U.N.S.W.
Paul McNeill, Community Medicine, U.N.S.W.
Christopher Magarey, School of Surgery, U.N.S.W.
John Powell, T.E.R.C., U.N.S.W.
Tim Sowerbutts

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

J.P. Powell

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

July 1978

Tertiary Education Research Centre
University of New South Wales

IR007808

ABSTRACT

A method for measuring the interpersonal communication skills of medical students was developed and evaluated. The stimulus material consisted of short films portraying patients expressing strong emotions to doctors and the students were required to make written responses indicating what they would do or say if they were the doctor. The test was trialled with first year medical students who were taking a course in interpersonal communication skills and a control group of first year physics students.

A rating schedule was developed which assessed ability to respond facilitatively in a patient-doctor interaction which the course was designed to encourage. Both open-ended and forced-choice responses were rated. The medical students' scores increased significantly from the pre-test to the post-test, in both the open-ended and forced-choice formats, while those of the physics students did not, indicating that the test was sensitive to changes in the students' interpersonal communication skills.

1. Introduction.

The growing interest in teaching interpersonal skills to those in helping professions has led to the demand for instruments to assess the efficacy of programmes that have been developed. Particularly this is so where the value of 'human relations' style programmes is seen to be in doubt if the effectiveness cannot be demonstrated. The following paper reports on one attempt to develop an instrument to assess interpersonal communication skills. The instrument was developed in a medical context but the procedures and material reported on could be modified to suit other professional contexts.

The test developed consists of a series of filmed vignettes, that is short scenes of patient-doctor interactions during which the patient expresses strong emotion. The vignettes end abruptly before the doctor replies and the testee is then asked to respond to the patient either in writing (open-ended test) or by choosing the most appropriate of five given responses (forced-choice test).

The test was developed initially to help assess the effectiveness of a new course Introductory Clinical Studies (ICS) which was introduced in the Medical Faculty of the University of New South Wales for all first year medical students in 1974. This course was designed to encourage students to be aware of the complexities of human relationships in the practice of medicine, including their own relationships with others.

A traditional approach to doctor-patient relations has been described as follows:

Because he is confronted with a succession of clients, many of whom are in some distress, a doctor must not allow himself to become too emotionally involved in any one case. To do so may impair his technical judgement. Furthermore, because he sees a large number of patients in a day, he must be able to "switch off" after seeing one in order to be ready for the next. This professional demeanour of the medical man is sometimes misread as cynicism; in fact, it is more accurately described as a professional detachment which enables him to apply specialised technical skills to the problem in hand with the greatest efficiency.¹

ICS was introduced to encourage a different approach. The objectives of the course included the expectations that the students would be better able to analyse the processes of communication that they observed, to communicate effectively with other people, to understand the ways in which they affected other people, and to respond appropriately to the expression of emotion.²

The test focuses on the testee's ability to respond appropriately

to expressions of emotion in doctor-patient interactions. 'Appropriately' means a style of response which recognises the existence of feelings and encourages their expression. The rationale for this is that satisfactory communication of content (fact and ideas) is unlikely to take place in the presence of emotional stress and that an open, honest response is likely to result in the subsequent reduction in stress. The test therefore is intended to give an indication of the extent to which someone has the capacity to display this style of responding - a style that may be termed facilitative.

According to Rogers (1961)³ significant learning is facilitated when the therapist exhibits congruence (genuineness), unconditional positive regard and accurate empathy. Important too is respect for the client as an individual. The client has the right to self-direction. We need to ask "Do we tend to treat individuals as persons of worth, or do we subtly devalue them by our attitudes and behaviour?"⁴ These dimensions are basic to client-centered therapy and to any facilitative interpersonal communication for those in the helping professions. Rogers suggests that traditional professionalism is developed as a defense because we are afraid to let ourselves freely experience positive feelings.⁵

Such an approach as advocated by Rogers was accepted as necessary for the effective functioning of the doctor as a helping professional in the development of this test. The test can be said to operationalise those dimensions of facilitative interpersonal communication that are significant for medical interview as distinct from a client-therapist encounter.

2. The Teaching and Assessing of Interpersonal Communication Skills.

2.1. The Use of Film or Video Material

The proliferation of videotape recorders in recent years has led to their increasing use in teaching at the tertiary level. They have proved useful tools in many professional programmes where effective interpersonal interaction is important. Rather than learning on the job, the student can practise skills with simulated clients by looking at videoed role plays or by viewing controlled encounters with actual clients. In teaching, the prospective teacher practises necessary skills one at a time in microteaching programmes.⁶ In psychiatry videoed material has been used extensively to

improve both patients' self-awareness and the skills of therapists.⁷

In general medicine videotape recorders have proved a useful adjunct in the increasing number of programmes that have aimed to teach interviewing skills and, more broadly, interpersonal skills, in the USA, Britain and Canada. The growth of such programmes is based on an awareness that a poor interview will not lead to an accurate diagnosis, even of organic disorders. Nor will a patient be likely to follow through with after care where no rapport is established.⁸

Many of these new programmes are structured around the use of videotape recorders to replay to the student interviews conducted by themselves or others, followed by small group discussion. On video the students can see all the communication activities, verbal and nonverbal, that are so important in an interview.⁹ However, the way in which the interviewing skills are taught, with or without the use of video or film material, varies and the effectiveness of the different techniques is not clearly established. Attempts to assess such programmes have produced conflicting results. Some of the problems involved are discussed in a brief review of the literature on teaching interviewing by Ward and Stein.¹⁰

2.2 Assessment Using Video or Film Material

Where video or filmed material is being used to teach interviewing or interpersonal skills to students, efforts have been made to test the students' learning with such material. One approach is to video a student with a real or simulated patient and have that interview rated by a group of judges according to some rating schedule.¹¹

Another approach is to have films of interviews that are stopped at crucial points at which the students choose, from a given set of alternatives, the best response to the patient in that situation.¹² One example of this type of test is that of Enelow, Adler and Wexler.¹³ This group developed a series (Programmed Instruction in Medical Interviewing) of films of interviews which are stopped, choices presented and the consequences of the different choices shown. In the test situation the interview is stopped nine times and three alternatives given, but the consequences are not shown.

A further example of the use of video or film for testing is the instrument developed by Kagan and associates in conjunction with their work on Interpersonal Process Recall. This instrument is the Affective Sensitivity Scale which has been tested as a predictor of empathy skills and used to measure change in sensitivity as a result of training experiences. For the Affective Sensitivity

Scale, the examinees watch a series of short vignettes and after each they respond in a forced choice format, selecting responses that are felt most likely to be what one of the participants in the encounter is really saying to himself at the end of the scene.¹⁴

There are problems associated with all of the methods of assessment discussed so far. Particularly difficult is determining how successful such tests are as predictors of a student's performance in a real life situation. It is possible for students to be able to be perceptive in discussing taped interviews and interview skills, but not perform well themselves.¹⁵ Kagan allows that while affective sensitivity as he defines and tests it, is a necessary precondition for being an effective counsellor, it is possible for someone with a high ranking on the scale to be an indifferent counsellor.¹⁶

The preceding discussion and associated references deal with the psychiatric interview, the general medical interview and that of the counsellor. The skills referred to range from the more prosaic ones of history taking to establishing empathy. No effort has been made to separate out the various strands as the types of interviews and skills overlap a great deal depending on the people involved, the type of programme and the view of the function of a doctor assumed to be desirable.

2.3 Situational Testing for Facilitative Interpersonal Communication

Situational tests do not necessarily involve film or video material. Some use audio-taped or written stimulus material. A test that is close to the one being reported here is the Physician-Patient Situation Test.¹⁷ This test consists of 35 typical physician-patient incidents in which a patient's comment to the physician and five possible responses are presented. The five possible responses represent five categories that are deemed independent. The categories are: evaluative, hostile, reassuring, probing and understanding, the last being seen as the most desirable type of response. Results from the use of this test show that students are better able to differentiate between responses than to use them in an interview, which accords with the results of other studies. At the end of the course, however, the students did show a grasp of a general concept of facilitative responding.

The Patient Situation Test, like the test being discussed here is an indirect measure. Both are assessing the ability to use a facilitative style of approach to patients or at least recognize the need to use such an approach. Neither the ability to respond in this way nor to recognise such a response is something that a student is likely to have without prior training of some sort.

Support for this contention comes from a study carried out by Phillips and Agnew to test Rogers' "Reflection" hypothesis. Using Rogers' definition of reflection as a type of response where the counsellor "...is trying to understand from the client's point of view and to communicate that understanding", Phillips and Agnew asked trained and untrained people to choose responses to a client in a hypothetical counselling situation. The results demonstrated that the untrained people did respond primarily in terms of non-reflective choices (probing, interpretation, support, etc.) while the trained people, though not necessarily adherents of the Rogerian school of therapy, responded more often in the reflective mode. Phillips and Agnew concluded that the skills being tested were "more than a simple extension of knowledge of interpersonal relations possessed by any reasonably intelligent and emotionally mature person".¹⁸

Further support for a facilitative style of responding being a learned skill comes from the work of Carkhuff and his associates. Carkhuff also distinguishes between the ability to discriminate facilitative dimensions and to communicate these dimensions. But he claims both can be learned, and in his programme for training counsellors and other helping professionals he sees the need to concentrate on both didactic and experiential learning.¹⁹

Although it was not used by the developers of the test this paper reports on, the work of Carkhuff is parallel in many areas. For testing he uses stimulus material to which testees respond; but the material is audio-taped. The responses are rated on varying scales of facilitative dimensions. His scale for rating gross levels of interpersonal functioning describes a facilitator in a way that accords closely with the interpersonal functioning deemed desirable and incorporated in our rating scale.

The facilitator is a person who is living effectively himself and who discloses himself in a genuine and constructive fashion in response to others. He communicates an accurate empathic understanding and a respect for all of the feelings of other persons and guides discussions with those persons into specific feelings and experiences. He communicates confidence in what he is doing and is spontaneous and intense. In addition, while he is open and flexible in his relationships with others, in his commitment to the welfare of the other person he is quite capable of active, assertive, and even confronting behaviour when it is appropriate.²⁰

3. The Development of the Test 1975 and 1976

The development in 1975 of a test using vignettes at the University of New South Wales was considerably influenced by the Kagan materials. They were being used for teaching purposes at the university but found not completely satisfactory for use with Australian students. The Tertiary Education Research

Centre was simultaneously developing the use of vignettes for staff development purposes. From these sources came the notion of developing a test of the interpersonal skills of ICS students using their responses to video, and later film, vignettes. A subsidiary goal was to produce film material for teaching purposes.²¹

3.1 The Film Stimulus Material

The film material for the test consists of short vignettes of doctor-patient interactions which last from 2-3 minutes. They are produced in colour on 16mm film for durability. Both for the test and for teaching sequences there are no titles or introductions to bias the responses of the viewers.

The filmed vignettes were made in a series of stages. First, ideas of suitable encounters were roleplayed by the research team and helpers and videoed. Those considered satisfactory as stimulus material were then audio-taped to provide a script. The script and the videos were given to trained actors who then developed the situations further before a film crew. The method used produced vignettes that were spontaneous, alive and with a feel of authenticity. One of the research team too, was a medical practitioner who checked the videos and films for medical details.

However the latitude given to the roleplayers and actors sometimes led to inappropriate nonverbal behaviour that was almost involuntary but marred the vignette. For example an attempt was made to film a male doctor and male patient where the patient expresses gratitude by touching the Doctor. This is contrary to cultural norms and proved very difficult. No completely satisfactory vignette was made. On the other hand unprompted extra meaning sometimes made vignettes richer. Two that are now part of a group of vignettes for teaching purposes, have a middle-aged man visiting the doctor soon after his wife's death from cancer. He thanks the doctor for all he's done for the family. These two vignettes of the same situation have unintended different meanings. In one the man does not look at the doctor and projects strongly the nonverbal message of hostility to the doctor. In the other with eye contact the message (intended) of gratitude is conveyed.

Guidelines for making suitable vignettes for teaching and testing were arrived at out of our experiences with using the vignettes and rating responses in 1976. The following criteria were agreed upon:

1. The content in the mind of the doctor must be clear, e.g. by some preliminary discussion with another figure or by some other preliminary activity.
2. The need for the doctor to deal with affect must be clear, e.g. the affect must be seen to be getting in the way of dealing with content or the affect makes termination of the encounter difficult.
3. The doctor must remain neutral and uncommitted in dealing with affect or content in any particular way so that the respondent (viewer, examinee) has no cues as to any particular response. In particular, the key figure must avoid any obviously inappropriate behaviour or comment.
4. The vignette must start slowly so that viewers are able to orientate themselves.
5. The vignette must finish abruptly with as little interaction as is necessary for the patient to indicate their affect and concern.

During 1976 new vignettes were made to satisfy these criteria and incorporated in the test film for 1977.²² The development of the new film material in 1976 was a reversal of the procedure originally applied in 1975. The early rating schedules were devised to rate responses to existing vignettes whereas now vignettes were being made to meet the needs of the rating schedule. The schedule for rating developed in early 1976 demanded vignettes where the need for an affective response was so clear that purely informational, cognitive responses could not be deemed at all appropriate.

Such vignettes needed to be very unambiguous. The subject, in order to respond as the doctor, needed to know the facts of the case so that the giving of alternative versions of the case would clearly indicate a desire to avoid confronting the emotions involved. For example, one vignette made in 1975 showed a doctor coming in to talk to a young man whose mother had just been operated on. The man asks, "She's alright isn't she?" The doctor does not reply - end of scene. Many students responded by telling the man that his mother would recover. Is that avoidance of discussing death, false reassurance, or a genuine misinterpretation? Such ambiguity had to be removed before the responses could be rated satisfactorily.

It was this fresh look at the film material which finally made clear that the type of material suitable for teaching purposes was not always suitable for testing. Ambiguity makes vignettes potent teaching material but poor test material. This conclusion partially accords with that of Levit in a different but related area. She produced films for teaching and testing clinical diagnosis and states that "the testing film differed fundamentally in both its objectives and its approach from the film produced for teaching purposes."²³ Our experience suggests that test material can be used for teaching, but teaching material is not necessarily appropriate for testing.

3.2 Development of the Rating Schedule Response Sheet

Following preliminary work in 1975 in 1976 filmed vignettes of doctor-patient interactions were shown to students in Year 1 Medicine, their ICS tutors and to other groups who included a group of experts who were academics or in the helping professions (doctors, psychologists or social workers) with a reputation for being able to establish empathy with other people, a group of health professionals who had not been trained in interpersonal communication skills and a group of fourth year medical students after they had completed their courses in introductory medicine and surgery, which had included traditional instruction in communication with patients and history taking. They were asked to write down what they had observed, how they would have responded and why.

Various rating scales were developed to rate the different parts of these responses but difficulties were encountered with such elaborate schedules. Therefore it was decided in early 1976 to simplify the rating schedule. In each vignette the raters would rate the affective response of a respondent on a single scale; the other material was used to help interpret the response where appropriate. The rationale for this was that correct observation of the filmed patients' or students' behaviour, and awareness of the patients' or students' emotional needs, were prerequisites for responding appropriately. Furthermore, this allowed the questions and rating for each vignette to be standardised.

The difficulties of correctly interpreting the significance and implications of a short statement in direct speech led us to continue asking respondents to explain their response in the new format now being used. (Appendix 1)

3.3 The Rating Scale

The main question to be answered in developing the rating scale was what constituted an appropriate response. The research team developed guidelines for rating each vignette by looking at a sample of 50 response sheets, 24 of which were from tutors and 26 from the students at the beginning of 1976. Guidelines for rating the responses were based on the raters' perceptions of to what extent each response was facilitative in style and would encourage the open expression of emotion. These guidelines for rating the responses to each vignette were then summarized into a general set of criteria which gave five categories of responses forming a rating scale of facilitative responsiveness as follows:

1. Rejection of the patient or avoidance of the truth.
2. No empathy. No acknowledgement of the emotion or feelings expressed by the patient.
3. Empathy demonstrated. Feelings of the patient acknowledged (at least implicitly) but without any attempt to explore or allow vent to feelings.
4. The patient invited in some way to explore, express or give vent to his feelings, but with clear limits imposed on the type of feelings or their method of expression.
5. The patient encouraged to express whatever feelings he wants to express, even if this might lead to some difficulty or embarrassment for the doctor.

Each category is allocated a score of 1 to 5 and it is the sum of the score given by each rater that produces the respondent's score for a vignette.

3.4 Preliminary Analysis in 1976

The work on the rating scale, the format of the test and the film material, were in themselves satisfying results from the trial project in 1976; however, although not all the vignettes in the Test were considered effective after analysing a preliminary sample of 50, it was decided to rate all the responses from the various groups for the 3 vignettes that were considered effective to try out the new rating scale and to gain some indication for future developments.

A rudimentary analysis of the results was carried out. The results were encouraging. Those groups who were expected to do better, the tutors and experts, did so. There was a strong indication that ability to score

highly on the test did not correspond with maturity, the non-experts and fourth year students did not score highly. Thus the test seemed to have consistency, to be measuring differences in interpersonal skills and the rating procedure was workable.

4. The 1977 Programme

The 1977 programme was designed to investigate further the value of the materials and procedures used in 1976.

4.1 The Test Instrument 1977

The test consisted of six vignettes of which three were new and one was modified. The respondents, as before, were shown the vignettes one at a time and asked to respond on the sheet provided for free responses. Following this, the respondents were given a response sheet containing questions about biographical data and a place to respond to a repeat of three of the vignettes, but this time using a forced-choice format for each vignette.

The vignettes all dealt with doctor-patient relations and required the respondents to reply to patients expressing emotion.

Vignette 1 A doctor tells a female patient after an examination that she has not got breast cancer. The woman thanks him, holding his hand.

Vignette 2 A patient reacts angrily to a doctor when he is told that the tests show nothing is wrong with him.

Vignette 3 A patient is very concerned at the doctor's suggestion of an operation.

Vignette 4 The patient is unhappy with the doctor's suggestion of cortisone treatment and asks whether the doctor knows what he is doing.

Vignette 5 The doctor's diagnosis of scabies is rejected by the children's mother.

Vignette 6 A doctor has to tell a young man that his mother is likely to die after an operation.

4.2 Forced Choice Responses

It was decided for 1977 to test the feasibility of a forced-choice format. If such a format could be developed it would make the test standardised

and usable by other groups, and would greatly reduce the work of rating. It was also felt that the open-ended responses could be biased in favour of the tutors and experts who might have a better facility from experience, at writing appropriate responses.

The alternative responses in the forced-choice section were drawn from previous responses to the vignettes and represent each of the categories on the five-point scale for each vignette. The student chooses the alternative which he considers most appropriate and is scored accordingly from 1 to 5.

4.3 Procedure

1. The pre-test was administered to all first year medical students in groups of 60 over a four day period. These groups correspond to their tutorial days for ICS during the year. In 1977, following recommendations arising from the evaluator of the course in 1976, ICS was structured differently. Students met in groups of 60 one day of the week with a team of six tutors. This team of tutors and a consultant worked out the programme in advance for their group which could choose to meet as a group of 60 or, more usually, in small groups of 10 to 12. The result of this arrangement was that each group of 60 students was following quite different programmes so that the ICS experience was not a constant for all 248 students.
2. The pre-test was also administered to two tutorial groups of Physics I students. Physics I students were chosen as being reasonably like medical students in previous training and ability. The groups were offered a small payment on completion of the post-test in an attempt to eliminate volunteer bias.
3. The tutors also took the pre-test.
4. The post-test was administered to all the above groups in September 1977.

4.4 Administration of the Test

The test was administered to all groups by the same person: the Head Tutor for ICS. The students were told that it was not a test, and that any questions about it would be answered afterwards. At the pre-test, despite this introduction students tended to behave as if in an examination room. For the post-test the students were much more relaxed, some even flippant.

As well as giving a general introduction the Head Tutor introduced the open-ended and forced-choice section in turn, by reading the introduction on the response sheet, emphasising (by reiteration) that the responses were to be "your response to emotion". The students were given 3 mins. for each of the open-ended responses and 1 min. each for the forced-choice responses.

5. The Analysis of the Results in 1977.

The aims of the 1977 analysis were:

- a. To assess whether the change in interpersonal skills (as measured by the test instrument) of the ICS students was the same as would be expected to occur as the results of the maturing influence of a year at University.
- b. To test for any significant differences according to sex and age of students.
- c. To check for any significant differences among the four tutor groups of 60 medical students.
- d. To test the feasibility of the forced-choice format.
- e. To compare the effectiveness of the six vignettes in discriminating ability to respond facilitatively.

5.1 Description of the Sample.

There were 245 medical students (nearly all the 1977 intake for Medicine 1) who took the pre-test. Of these, 152 persisted to take the post-test in September 1977 (henceforth referred to as the trainees). The physics students were in two tutorial groups totalling 106 students. Of these 60 returned in September (henceforth known as the controls).

TABLE 1.

Numbers of Medical and Physics I
students who took the pre-test and the post-test.

Group	Persistees	Non Persistees	Total
Medicine	152 71.7%	93 66.9%	245 69.8%
Physics	60 28.3%	46 33.1%	106 30.2%
Total	212 100.0%	139 100.0%	351 100.0%

The sex and age of the trainees and controls is given in Table 2.

TABLE 2.

Sex and Age Characteristics of
Trainees and Controls.

Group	SEX		AGE		Total
	Male	Female	17 to 19	20 & over	
Medicine	110 69.6%	42 77.8%	112 67.1%	40 83.9%	152 71.7%
Physics	48 30.4%	12 22.2%	55 32.9%	5 11.1%	60 28.3%
Total	158 100.0%	54 100.0%	167 100.0%	45 100.0%	212 100.0%

There are more males than females in both groups which is a reflection of the original population taking the pre-test. Interestingly, more of the trainees are over the age of 19 years and must therefore have had some experience other than high school before beginning Medicine 1.

5.2. Rater Reliability

The open-ended responses to the pre- and post-tests were rated by three raters. To ensure that there was no bias, the response sheets for pre- and post-tests were coded and jumbled by an outsider. The raters then met six times, training for each scene before they then each rated the responses for that scene. The raters achieved a very satisfactory degree of reliability.

TABLE 3.

Reliability coefficients of the total score using the entire sample, including tutors (N = 374).

VIGNETTE.	PRE	POST
1	0.900	0.911
2	0.907	0.913
3	0.898	0.921
4	0.871	0.887
5	0.852	0.874
6	0.834	0.866

Note: All coefficients are statistically significant beyond the 0.001 level.

The statistics given in Table 3 are the reliabilities for the average rating of the three raters.²⁴ These reliabilities are intraclass correlations, which can be statistically interpreted as the average intercorrelations between the raters. Since the intercorrelation between any two raters always lies between +1 and -1 a value in the region of +0.9 is very high and it can be concluded that for these raters there is a great deal of reliability. Further the analysis was done using the average score over the three raters.

It is of interest to note that the reliability of Vignette 6 although high, is not quite so high as the other scenes - this point will be discussed later when an analysis of individual scenes is made.

5.3 Non-Completion of the Course.

Since a sizeable group of both the medical and physics students did not persist to take the post-test, the two groups, those who persisted and those who did not, were tested for any significant difference on the pre-test.

TABLE 4
Average score for the total of the six vignettes
on the pre-test (Open responses)

Groups.	Persistees.	Non Persistees.	Total
Medicine	14.08	14.08	14.08
Physics	12.44	12.91.	12.65.
Total	13.62.	13.62.	13.65.

The analysis of the total scores (the sum of the scores on the six vignettes on the pre-test shows there was no significant interaction between Persistence and Treatment (i.e. Medicine or Physics) ($F_{1,326} = 0.737$, $p = 0.391$) nor was there a significant difference between the persistees and the non-persistees ($F_{1, 326} = 0.210$, $p = 0.647$). Hence it can be concluded that in terms of the pre-test scores the persistees were not different from the non-persistees and this effect held for both the medical and physics students.

Although there was no difference between the persisters and non-persisters on the pre-test, we do not know if the non-persisters would have been subject to differential change rates. It was decided to use only the persisters for the analysis that follows as this allows a matched analysis to be used which is more efficient. The rest of the analysis will be done using only those who persisted (the trainees and the controls).

5.4 Statistical Analysis.

TABLE 5
Average total score for all six vignettes on the pre-test
for Trainees and Controls by age (Open response).

Groups.	17 to 19 years.	20 years & over.	Total
Trainee	14.21	13.71.	14.08.
Control	12.41	12.83.	12.44
Total	13.62.	13.62.	13.62.

In order to provide a summary statistic of the efficacy of the treatment the analysis was done on the total score, that is, the sum of the (average)

scores for each vignette. Later in the report the effects of the treatment on individual vignettes will be discussed. The statistical technique for assessing the effect of a treatment in a "before and after, controlled experiment" varies depending on whether the trainees and controls have significantly different scores on the pre-test. To this end the total score on the pre-test open response was submitted to an analysis of variance. This showed that there was no significant Age X Treatment interaction ($F_{1,193} = 0.428$, $p = 0.514$), no significant age effect ($F_{1,193} = 0.717$, $p = 0.396$) but a significant difference between the trainees and the controls ($F_{1,193} = 17.016$, $p = 0.001$).

From Table 5 it can be seen that the trainees have a significantly higher score than the controls on the pre-test and this difference is not accounted for by the different age distribution of the two samples.

Since the trainees and controls have different scores on the pre-test in order to test the efficacy of the ICS course it is necessary to look at the change of the trainees' average score from pre- to post-test and compare it with the corresponding change for the controls. The average total scores of the two groups on the pre-test (14.08 for trainees, 12.44 for the controls) are of the same order and a change of the order of 20% or less was expected (the actual change was less than 20%). Consequently an acceptable means of measuring the effect of the treatment is to compare the arithmetic (i.e. not the proportional) change in the average score of the trainees from pre- to post-test and corresponding change for the controls.

Table 6 gives an analysis of variance of the change of the average total score (per the open response).

TABLE 6

Analysis of variance of the change of the average total of open response.

Source of Variation	Degree of Freedom	Mean Square	F	Significance
Treatment *	1	88.924	14.748	0.001
Age	1	0.093	0.015	0.901
Sex	1	0.917	0.152	0.697
Age X Treatment	1	0.722	0.120	0.730
Sex X Treatment	1	0.366	0.061	0.806
Age X Sex	1	16.218	2.690	0.103
Error	168	6.029		
Total	174			

* 'Treatment' means whether the student was in the experimental (i.e. medical) group or the control (i.e. physics) group.

Since the Treatment X Age and Treatment X Sex interactions are non-significant the effect of the treatment does not appear to depend on the sex or age of the student. As a result the inference about the effect of the ICS course can be made using the total sample - as can be seen from Table 5 there is a highly significant ($p = 0.001$) difference between the change in the average total scores for the trainees and the controls. This difference was such that the trainee students increased the average total score from 14.08 to 15.66 while the control students only changed from 12.44 to 12.54.

5.5 Tutor Groups

In 1977 the ICS programme was structured so that the tutors who taught the same day of the week worked as a team in planning the course for the group of 60 students who met on the same day. It was thought that this might lead to significant differences among the groups. A one-way analysis of variance showed no significant difference in rates of change among the groups ($F_{3,124} = 1.44, p = 0.23$).

5.6 Individual Vignettes

Although using the average scores was the most efficacious way of looking at the change from pre-test to post-test, there is value in looking at the changes for individual vignettes.

For all vignettes the trainees' average score increased, unlike the average score for the controls and this accounts for the overall difference in change rates. In the average total scores the controls had a very small overall total increase made up of some increases in average scores and some decreases in average scores on individual vignettes.

The vignettes that differentiated most strongly between trainees and controls were 3 and 6, and less strongly 4. In vignette 3 and vignette 6 the trainees increased their average score while the controls actually had a decrease in their average score. In vignette 4 both trainees and controls increased their average scores but at different rates.

Average scores for individual vignettes, (Open Responses).

Group	Pre	Post	Increase	Probability of average change being the same for Trainee and Control students.
Trainee Vignette 1 Control	2.42	2.62	+0.20	$p = 0.697$
	1.98	2.20	+0.14	
Trainee Vignette 2 Control	2.27	2.41	+0.14	$p = 0.603$
	1.71	1.78	+0.07	
Trainee Vignette 3 Control	2.08	2.37	+0.29	$p = 0.002$
	1.85	1.73	-0.12	
Trainee Vignette 4 Control	2.15	2.45	+0.30	$p = 0.052$
	1.74	1.81	+0.07	
Trainee Vignette 5 Control	2.36	2.54	+0.18	$p = 0.207$
	2.24	2.26	+0.02	
Trainee Vignette 6 Control	2.73	3.12	+0.39	$p = 0.006$
	2.80	2.68	-0.12	

5.7. Open and Forced-Choice Format.

The purpose of using the forced-choice and open-ended format on the same vignette was to test the feasibility of the forced-choice format. The change in average scores from pre-test to post-test was in the same direction as on the open-ended section but the significance of the change was not quite so high. This is mainly due to the different distribution of the scores for the forced-choice items.

TABLE 8

Average scores for individual vignettes, (Forced-choice responses)

Group	Pre	Post	Increase	Probability of average change being the same for Trainee and Control students.
Vignette 4	Trainee	2.73	3.01	p = 0.155
	Control	2.37	2.39	
Vignette 5	Trainee	3.10	3.25	p = 0.854
	Control	2.75	2.88	
Vignette 6	Trainee	2.85	3.13	p = 0.018
	Control	2.88	2.63	

By comparing Tables 7 and 8 it can be seen that the initial average scores for the forced-choice items were higher than for the open responses. This supported the initial hypothesis that students performed less well than tutors on the open-ended test through lack of written verbal ability.

Results from vignette 6 were conspicuously different. The trainees and controls had almost identical average scores on the pre-test of the forced-choice section but the controls then had a negative change, while the trainees increased their average score significantly.

6. Summary.

- a. From the statistical analysis of the 1977 results it can be seen that the trainee students did make significantly higher gains in average total scores as compared to the controls. The average magnitude of the change for each vignette for the trainee is +0.26. If we translate the scores according to the rating scale, we can say that the trainees began at +2.35 which is slightly above category 2 where no empathy is demonstrated and improved to +2.61 which is closer to category 3 where empathy is demonstrated,

though without any effort to explore the patients' feelings. This change although small is significant. The results need to be seen in context. ICS is not a course in "How to do Human Relations". At no point is there any effort to teach the test or even, directly and explicitly, the mode of responding to patients that is allocated the highest scores on the rating scale.

Overall the analysis bears out that the change in interpersonal skills is more than would be expected from the result of spending a year at University.

- b. The results showed no significant differences according to sex and age of students which bore out earlier suggestions that the test measures something that is not necessarily related to maturity.
- c. There were no significant differences among the four tutor groups. If as Carkhuff suggests, there is a significant relationship between level of therapist functioning and change in helpee²⁵ this would not show up in a test of change over a group of tutors and their students. Nor might it, even with more detailed analysis, as the ICS tutors are not functioning as therapists.
- d. The results of the analysis showed that the forced-choice format is feasible. It has been demonstrated that the forced-choice format does measure the same changes as the open-ended format and can be used in its place. A forced-choice test would be a more conservative instrument in that changes of scores are of a lower magnitude but this would be outweighed by the advantages in administering the test and processing the results. The open-ended format would still be very valuable for use with small numbers for discussion, self assessment or diagnostic work by the teacher. For large numbers, it is not feasible. Training raters would be a major job in itself and the subsequent rating of responses would take an unacceptable length of time. It took each rater on this project about 6 hours to rate 563 responses for one vignette. The forced-choice format could be processed manually or modified for machine scoring to allow easy and economic processing.

- e. From a comparison of the results for the six vignettes it appears that vignettes 3, 6 and to a lesser extent 4 differentiated most strongly between trainees and controls. The explanation for these results could lie in the nature of the vignettes. Vignettes 3 and 6 relate to hospitals and show scenes where knowing what to say would most likely have been touched on in the ICS programme. In scenes 2, 4 and 5 there is a challenge to the doctor's authority that might lead to the controls identifying somewhat with the patient rather than the doctor and this could influence scores.

Vignette 6 was thought by its producers to have the strongest emotional impact and it certainly produced interesting results. Rater reliability was lowest (though still sufficient). The vignette differentiated most strongly between trainee and controls on the open-ended responses and likewise on the forced-choice responses (see Table 8) where the controls' average score again decreased.

It was encouraging that vignettes 3 and 6 were made according to the guidelines produced in 1976. However more work needs to be done on producing satisfactory vignettes, especially as the test in the forced-choice format would need to be longer.

7. Other Uses of the Materials Developed

While the main emphasis of this report has been on the development of testing procedures the filmed vignettes have other uses.

7.1 Teaching Materials

Some of the vignettes produced in 1976 have been made available for teaching purposes by the Tertiary Education Research Centre in a series of Trigger Films on Doctor-patient relations. Teaching notes compiled by Susan Hayes, Department of Behavioural Sciences in Medicine, University of Sydney, accompany the series. The series includes three groups: 'First Meeting', 'The Doctor an Expert' and 'Communication Problems'.

Dr. Hayes sets out the advantages of the series for teaching health professionals as follows:

The episodes in this film all involve situations in which there is some form of real or potential difficulty in the patient-doctor relationship. Some of them are strongly emotional in tone and this has deliberately been made explicit, even if this involves a degree of exaggeration, in order to provide a basis for discussion. Although it may be unusual for patients and doctors openly to show their emotions there is no doubt that they would often like to be able to do so. What we see here is the open expression of feelings, reactions, and difficulties which are often kept hidden and only rarely allowed to surface.

Doctor-patient interaction involves a situation where emotional and factual content are irrevocably intermingled. Each person in the interaction responds to the other on rational and irrational levels - it is important for health professionals to be able to recognise what the patient may be experiencing, as well as identifying their own responses.²⁶

The use of the open-ended response sheet in teaching has already been mentioned. The rating scale of categories could likewise be useful in teaching. The open responses reveal a great deal about student attitudes and discussion of these attitudes identifying them according to the rating scale categories could be productive.

A set of vignettes depicting simulated nursing situations were also developed at the Tertiary Education Research Centre. Some of these vignettes were used in the 1976 test project; all were made according to the techniques developed during this project. These vignettes have also been made available for teaching purposes by T.E.R.C.

7.2 Evaluation

Another use for such vignettes is evaluation. The nursing vignettes were used in 1976 as part of an evaluation of nursing programmes in the Australian Capital Territory. The responses of students to the vignettes were analysed to give information about the professional attitudes being developed by the different nursing programmes.²⁷

The doctor-patient vignettes which have forced-choice alternatives available have been used experimentally by the Faculty of Medicine at Newcastle as part of their attempt to develop an alternative method of selecting medical students.

8. Discussion

A number of issues arise from this study which merit further discussion and the most important of these seem to be the following: What is the test measuring? Why are some film scenes more effective than others? Is a

forced-choice format feasible? Are there other applications for the material? Do the results justify further development work?

1. Because the test had its origins in the need to assess student progress in a course dealing with interpersonal communication skills the main concern was to focus on the ability of students to respond appropriately to expressions of emotion in doctor-patient interactions. As was discussed earlier the test was designed to indicate ability to respond facilitatively in such a situation.

The test is not just measuring the ability to give the responses which students might believe are wanted: there are no cues in the written or visual material to enable students to "read off" the appropriate response. Nor is the appropriate response obvious from general experience: cultural norms would probably encourage low-scoring responses. The value system which is built into the test would probably not be widely accepted in the community. Thus their value needs to be accepted and the appropriate style of response needs to be learnt. Finally, as was pointed out earlier in the report, the test is not measuring something which could be expected necessarily to be acquired along with increasing maturity.

It is recognized that the predictive power of the test is limited with respect to performance in real-life situations. However, the formulation or recognition of an appropriate response (tested respectively by the open-ended and forced-choice formats) must precede its being spoken so the scores obtained reflect the potential for responding in a real situation.

2. The content of the film vignettes was obviously of crucial importance and in the early stages it was far from clear what general principles should be used to guide their production. It became evident later, after a good deal of trial and error, that (a) emotions should be expressed clearly and unambiguously, (b) the factual context from which the expression of feelings arose should be quite explicit and (c) there should be minimal focus of the doctor with the vignette ending prior to his or her response so that cues were as far as possible eliminated. Relatively few of the film vignettes made satisfied all of these criteria and most were therefore unsuitable for use in the test.

Of the six vignettes selected for the tests in 1977, in the belief that fulfilled the above criteria, only three satisfactorily distinguished between

the medical students who undertook training in interpersonal skills and the physics students who did not. Each of these showed the doctor as the active treater of disease. In one the doctor was a surgeon just after an operation, in another the doctor was advising the patient to undergo an operation and in the third the doctor was prescribing cortisone for a skin condition. The role of the doctor in the other three vignettes was less active: in one the female patient thanks him and touches his hand, in another the patient expresses anger at the doctor's diagnosis and in the third the patient is indignant at the possible suggestion of psychosomatic illness. While the medical students' scores increased in each case, the increase was greater with the former three and was not significantly greater than the physics students with the latter three.

This suggests that the context of the vignettes is as important as their content. The medical students probably focussed their attention during the year on the role of the doctor and learnt communication skills in the context of the doctor being the active treater of disease. Other roles of the doctor such as comforter and friend are far less frequently adopted and unlikely to have been experienced by the students, so they were less able to apply their newly learnt skills in the unfamiliar context. If this type of test is to be used in other professional training programmes it may be necessary to produce material specific to each professional setting.

3. It was hoped eventually to produce a test that could be widely used with large numbers of students. The use of an open-ended format which involved the scoring of thousands of responses made the need for a forced-choice format very clear. So it was pleasing in 1977 to find that the forced-choice format was feasible and that this type of test could be administered economically without loss of effectiveness to large numbers of subjects. This would eliminate the need to train raters to undertake a complex task and also permit the responses to be machine-scored.

4. The film material produced during this project constitutes a valuable resource for teaching purposes and some of it is already being used in medical and nursing education programmes. The open-ended format of the test can also be used by teachers to provide information on student attitudes. The test is currently being used experimentally as part of the selection procedure for entry to the medical school at the University of Newcastle. Additionally, this type of test could make a contribution to the evaluation programmes designed to promote moral and affective change. There seems to

be no reason why tests of this form could not be developed for use in areas such as management, personnel, social work and teaching.

5. The results of this project appear sufficiently encouraging to warrant further work in this area especially in view of the fact that it has been very much neglected by researchers. There is also evidence of a growing interest in the role of the emotions in professional training and in education generally. This will inevitably increase the demand for suitable assessment and evaluation instruments.

ACKNOWLEDGMENTS.

Work on this project was assisted by a grant from the Education Research and Development Committee.

REFERENCES:

1. Anderson, D.S. & Western, J. (1976) "The Professions: Reason and Rhetoric" in Boreham et al., The Professions in Australia: A Critical Appraisal, Brisbane: University of Queensland Press, p. 49.
2. Magarey, C.J. "Evaluation of Introductory Clinical Studies (ICS) in 1976", Unpublished Report, University of N.S.W., p. 3.
3. Rogers, C.R. (1961) On Becoming a Person, Boston: Houghton Mifflin Co., especially pp. 279-295.
4. Rogers, C.R. (1965) Client-Centred Therapy, Boston: Houghton Mifflin Co., p. 20.
5. Rogers, C.R. (1961) op. cit., p. 52.
6. Clift, J.C. & Malley, J.I. (1974) Bibliography of Microteaching with Selected Annotations, Wellington, N.Z.: University Teaching and Research Centre, Victoria University.
7. Kenny, B.L. (1969) Annotated Bibliography on Television and Videotape in Psychiatry, Jackson, Miss: Department of Psychiatry, University of Mississippi Medical Centre.
8. Maguire, P. & Rutter, D. (1976) "Training Medical Students to Communicate" in Bennett, A.E. ed., Communication Between Doctors and Patients, London: Oxford University Press, pp. 47-74.
9. See for examples: Meadow, R. & Hewitt, C. (1972) "Teaching Communication Skills with the Help of Actresses and Video-tape Simulation", British Journal of Medical Education, 6: 317-22; Secundy, M.G. & Katz, V. (1975) "Factors in Patient/Doctor Communication: A Communication Skills Elective", Journal of Medical Education, 50: 689-91; Waldron, J. (1975) "Teaching Communication Skills in Medical School", American Journal of Psychiatry, 130: 578-81.
10. Ward, N.G. & Stein, L. (1975) "Reducing Emotional Distance: A New Method to

Teach Interviewing Skills", Journal of Medical Education, 50: 605-13

11. See for examples: Anderson, J. et al. (1970) "The Definition and Evaluation of the Skills Required to Obtain a Patient's History of illness: The use of Videotape recordings". Postgraduate Medical Journal, 46: 606-12; Barbee, R.A. et al. (1967) "The Quantitative Evaluation of Student Performance in the Medical Interview", Journal of Medical Education, 43: 238-43; Goroll, A.H. et al. (1974) "Teaching the Clinical Interview: An Experiment with First-Year Students", Journal of Medical Education, 49: 851-67; Jarrett, F.J. et al. (1972) "Measuring Interviewing Skill", Canadian Psychiatric Association Journal, 17: 183-88; Turner, E.V. et al. (1974) "Predictors of Clinical Performance", Journal of Medical Education, 49: 338-42.
12. Muslin, H.L. et al. (1968) "Research in Medical Education: Research and Learning in Psychiatric Interviewing", Journal of Medical Education, 43: 398-404.
13. Enelow, A.J. et al. (1970) "Programmed instruction in Interviewing. An Experiment in Medical Education", Journal of the American Medical Association, 212: 1843-46.
14. Kagan, N. & Krathwohl, D.R. (1967) Studies in Human Interaction, East Lansing: Educational Publication Services, Michigan State University; Danish, S.J. and Kagan, N. (1971) "Measurement of Affective Sensitivity: Towards a Valid Measure of Interpersonal Perception", Journal of Counselling Psychology, 18: 51-54.
15. Jarrett, F.J. et al. (1972) op. cit.; Ware, J.E. et al. (1971) "A Negative Relationship Between Understanding Interviewing Principles and Interview Performance", Journal of Medical Education, 46: 620-22.
16. Kagan, N. and Krathwohl, D.R. (1967) op. cit.
17. Rasche, L.M. et al. (1974) "Evaluation of a Systematic Approach to Teaching Interviewing", Journal of Medical Education, 49: 589-595.
18. Phillips, E.L. and Agnew, J.W. (1953) "A Study of Rogers 'Reflection' Hypothesis", Journal of Clinical Psychology, 9: 281-84.

19. Carkhuff, R.R. (1969) Helping and Human Relations: A Primer for Lay and Professional Helpers, Vol 1, New York: Holt, Rinehart & Winston, Inc, passim.
20. Carkhuff, R. R. (1969) "Helper Communication as a Function of Helpee Affect and Content", Journal of Counselling Psychology, 16: 126-131
21. For further detail about the 1975 developments see:
Pearson, M. et. al (1977) The Development of Film Material and Testing Procedures to Assess Student Interpersonal Skills, Research and Development Paper No. 50. Sydney: T.E.R.C., U.N.S.W.
22. It is worth comparing these criteria with the guidelines given in Fisch, A.L. (1972) "The Trigger Film Technique", Improving College and University Teaching, 20: 286-89.
23. Levit, E.J. (1969) "Testing Physicians with Motion Pictures", Industrial Photography, 18: 20-2, 61-65.
24. Ebel, R.L. (1951) "Estimation of the Reliability of Ratings", Psychometrika, 16: 407-24.
25. Carkhuff R. R. (1969) "Critical Variables in Effective Counselor Training", Journal of Counselling Psychology, 16: 238-245.
26. Hayes, S. (1977) "Teaching Notes for Trigger Films:. Series 4", T.E.R.C., U.N.S.W., p. 3.
27. Brooker, R. et al.: "Evaluation of Basic Nursing Programmes in Australian Capital Territory: Students response to simulated Nursing situations", Sydney: Unisearch, U.N.S.W.

APPENDIX 1.

OPEN-ENDED RESPONSE SHEET 1977

--	--	--	--

DOCTOR PATIENT COMMUNICATION FILM

This exercise, which will take a little over half an hour, is designed to assess your ability to respond, as a doctor, to emotion expressed by a patient. You will be shown, in turn, six short scenes of a doctor and a patient and the film will be stopped after each scene.

Imagine that YOU ARE THE DOCTOR. Write down what you would do and/or say to the patient at this point. Be specific if you can and give reasons for your responses.

SCENE 1. Male doctor and female patient with breast symptom.

SCENE 2. Male doctor and male patient with abdominal pain.

SCENE 3. Female doctor and female patient with gall stones.

SCENE 4. Male doctor and male patient. Cortisone. (This scene is very short.)

SCENE 5. Female doctor, mother and two children with scabies. (This scene is very short.)

SCENE 6. Male surgeon, nurse and the son of a dying patient.
